THE ACM DIGITAL LIBRARY

Feedback

searching keywords jaccard distance
Terms used: searching keywords jaccard distance

Sort results by relevance

Display results expanded form

Save results to a Binder

□ Open results in a new window

Refine these results w
Try this search in The

Results 1 - 20 of 58

Result page: 1 2 3 next >>

1 LSH forest: self-tuning indexes for similarity search
Mayank Bawa, Tyson Condie, Prasanna Ganesan
May 2005 WWW '05: Proceedings of the 14th international conference on World Wide Web
Publisher: ACM
Full text available: pdf(247.91 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 10, Downloads (12 Months): 109, Citation Count: 1

We consider the problem of indexing high-dimensional data for answering (approximate) similar search queries. Similarity indexes prove to be important in a wide variety of settings: Web searc engines desire fast, parallel, main-memory-based indexes ...

Keywords: peer-to-peer (P2P), similarity indexes

2 Extracting redundancy-aware top-k patterns
Dong Xin, Hong Cheng, Xifeng Yan, Jiawei Han
August 2006 KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledg discovery and data mining
Publisher: ACM
Full text available: pdf(830.88 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 16, Downloads (12 Months): 125, Citation Count: 0

Observed in many applications, there is a potential need of extracting a small set of frequent patterns having not only high significance but also low redundancy. The significance is usually defined by the context of applications. Previous studies have ...

Keywords: pattern extraction, redundancy, significance

3 On synopses for distinct-value estimation under multiset operations
Kevin Beyer, Peter J. Haas, Berthold Reinwald, Yannis Sismanis, Rainer Gemulla
June 2007 SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data
Publisher: ACM
Full text available: pdf(309.19 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 21, Downloads (12 Months): 258, Citation Count: 0

The task of estimating the number of distinct values (DVs) in a large dataset arises in a wide va of settings in computer science and elsewhere. We provide DV estimation techniques that are designed for use within a flexible and scalable "synopsis ...

Keywords: distinct-value estimation, synopsis warehouse

4 VISTO: visual storyboard for web video browsing

Marco Furini, Filippo Geraci, Manuela Montangero, Marco Pellegrini

July 2007 CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval

Publisher: ACM

Full text available: pdf(398.68 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 1, Downloads (12 Months): 99, Citation Count: 0

Web video browsing is rapidly becoming a very popular activity in the Web scenario, causing the production of a concise video content representation a real need. Currently, static video summa techniques can be used to this aim. Unfortunately, they ...

Keywords: clustering, video browsing, video summary

5 Exploiting correlated keywords to improve approximate information filtering

Christian Zimmer, Christos Tryfonopoulos, Gerhard Weikum

July 2008 SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Rese and development in information retrieval

Publisher: ACM

Full text available: pdf(510.74 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 54, Downloads (12 Months): 0, Citation Count: 0

Information filtering, also referred to as publish/subscribe, complements one-time searching sin users are able to subscribe to information sources and be notified whenever new documents of interest are published. In approximate information filtering ...

Keywords: Peer-to-Peer (P2P), approximate publish/subscribe, distinct-value (DV) estimation, distributed information filtering (IF), information systems

6 Extraction and classification of dense communities in the web

Yon Dourisboure, Filippo Geraci, Marco Pellegrini

May 2007 WWW '07: Proceedings of the 16th international conference on World Wide Web

Publisher: ACM

Full text available: pdf(258.41 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 16, Downloads (12 Months): 261, Citation Count: 0

The World Wide Web (WWW) is rapidly becoming important for society as a medium for sharing data, information and services, and there is a growing interest in tools for understanding collect behaviors and emerging phenomena in the WWW. In this paper ...

Keywords: communities, dense subgraphs, web graph

7 From frequent itemsets to semantically meaningful visual patterns

Junsong Yuan, Ying Wu, Ming Yang

August 2007 KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledg discovery and data mining

Publisher: ACM

Full text available: pdf(2.09 MB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 51, Downloads (12 Months): 409, Citation Count: 0

Data mining techniques that are successful in transaction and text data may not be simply appli image data that contain high-dimensional features and have spatial structures. It is not a trivial to discover meaningful visual patterns in image ...

Keywords: image data mining, meaningful itemset mining, pattern summarization, self-superv clustering

8 SEPIA: estimating selectivities of approximate string predicates in large Databases
Liang Jin, Chen Li, Rares Vernica
August 2008 The VLDB Journal — The International Journal on Very Large Data Bases, Volu
Issue 5
Publisher: Springer-Verlag New York, Inc.
Additional Information: full citation, abstract

Bibliometrics: Downloads (6 Weeks): n/a, Downloads (12 Months): n/a, Citation Count: 0

Many database applications have the emerging need to support approximate queries that ask fo strings that are similar to a given string, such as "name similar to smith" and "telephone numbe similar to Keywords: Approximate, Estimation, SEPIA, Selectivity, String

9 THESUS: Organizing Web document collections based on link semantics
Maria Halkidi, Benjamin Nguyen, Iraklis Varlamis, Michalis Vazirgiannis
November 2003 The VLDB Journal — The International Journal on Very Large Data Bases, \
12 Issue 4
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(262.85 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 7, Downloads (12 Months): 86, Citation Count: 5

The requirements for effective search and management of the WWW are stronger than ever. Currently Web documents are classified based on their content not taking into account the fact t these documents are connected to each other by links. We claim ...

Keywords: Document clustering, Link analysis, Link management, Semantics, Similarity measu World Wide Web

10 Efficient similarity joins for near duplicate detection
Chuan Xiao, Wei Wang, Xuemin Lin, Jeffrey Xu Yu
April 2008 WWW '08: Proceeding of the 17th international conference on World Wide Web
Publisher: ACM
Full text available: pdf(327.62 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 39, Downloads (12 Months): 78, Citation Count: 0

With the increasing amount of data and the need to integrate data from multiple data sources, a challenging issue is to find near duplicate records efficiently. In this paper, we focus on efficient algorithms to find pairs of records such that their ...

Keywords: near duplicate detection, similarity join

11 On scalability of the similarity search in the world of peers
Michal Batko, David Novak, Fabrizio Falchi, Pavel Zezula
May 2006 InfoScale '06: Proceedings of the 1st international conference on Scalable information systems
Publisher: ACM
Full text available: pdf(297.44 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 3, Downloads (12 Months): 113, Citation Count: 3

Due to the increasing complexity of current digital data, similarity search has become a fundam computational task in many applications. Unfortunately, its costs are still high and the linear scalability of single server implementations prevents ...

12 Query clustering using user logs
January 2002 ACM Transactions on Information Systems (TOIS), Volume 20 Issue 1
Publisher: ACM
Full text available: pdf(1.31 MB) Additional Information: full citation, abstract, references, cited by, index term

review

Bibliometrics: Downloads (6 Weeks): 21, Downloads (12 Months): 176, Citation Count: 26

Query clustering is a process used to discover frequently asked questions or most popular topics a search engine. This process is crucial for search engines based on question-answering. Becaus the short lengths of queries, approaches based on ...

Keywords: Query clustering, search engine, user log, web data mining

13 Summarizing data using bottom-k sketches
Edith Cohen, Haim Kaplan
August 2007 PODC '07: Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing
Publisher: ACM
Full text available: pdf(269.48 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 4, Downloads (12 Months): 94, Citation Count: 0

A *Bottom-sketch* is a summary of a set of items with nonnegative weights that supports approxi query processing. A sketch is obtained by associating with each item in a ground set an indepen random rank drawn from a probability distribution ...

Keywords: all-distances sketches, bottom-k sketches, data streams

14 Comparative study of name disambiguation problem using a scalable blocking-based framework
Byung-Won On, Dongwon Lee, Jaewoo Kang, Prasenjit Mitra
June 2005 JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries
Publisher: ACM
Full text available: pdf(1.06 MB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 2, Downloads (12 Months): 60, Citation Count: 7

In this paper, we consider the problem of ambiguous author names in bibliographic citations, an comparatively study alternative approaches to identify and correct such name variants (e.g., "Vannevar Bush" and "V. Vush"). Our study is based on a scalable ...

Keywords: blocking, measuring distances, name disambiguation

15 Using web information for creating publication venue authority files
Denilson Alves Pereira, Berthier Ribeiro-Neto, Nivio Ziviani, Alberto H. F. Laender
June 2008 JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries
Publisher: ACM
Full text available: pdf(232.26 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 15, Downloads (12 Months): 11, Citation Count: 0

Citations to publication venues in the form of journal, conference and workshop contain spelling variants, acronyms, abbreviated forms and misspellings, all of which make more difficult to retri the item of interest. The task of discovering and reconciling ...

Keywords: authority file, bibliographic citation, canonical name, publication venue

16 Effective and scalable solutions for mixed and split citation problems in digital libraries
Dongwon Lee, Byung-Won On, Jaewoo Kang, Sanghyun Park
June 2005 IQIS '05: Proceedings of the 2nd international workshop on Information quality in information systems
Publisher: ACM
Full text available: pdf(695.11 KB) Additional Information: full citation, abstract, references, cited by

Bibliometrics: Downloads (6 Weeks): 4, Downloads (12 Months): 52, Citation Count: 4

In this paper, we consider two important problems that commonly occur in bibliographic digital libraries, which seriously degrade their data qualities: *Mixed Citation (MC)* problem (i.e., citatior different scholars with their names being homonyms ...

17 Localized signature table: fast similarity search on transaction data
Qiang Jing, Rui Yang, Panos Kalnis, Anthony K. H. Tung
November 2004 CIKM '04: Proceedings of the thirteenth ACM international conference on Informati and knowledge management
Publisher: ACM
Full text available: pdf(200.77 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 5, Downloads (12 Months): 49, Citation Count: 0

Recently, techniques for supporting efficient similarity search over huge transaction datasets ha emerged as an important research area. Several indexing schemes have been proposed towards direction. Typically, these schemes provide a tradeoff ...

Keywords: data mining, indexing, similarity search, transaction data

18 Interactive search of rules in medical data using multiobjective evolutionary algorithms
Daniela Zaharie, Diana Lungeanu, Flavia Zamfirache
July 2008 GECCO '08: Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation
Publisher: ACM
Full text available: pdf(368.37 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 19, Downloads (12 Months): 0, Citation Count: 0

In this work, we propose an approach for evolving rules from medical data based on an interacti multi-criteria evolutionary search: besides selecting the set of criteria and the sets of potential antecedent and consequent attributes, the user can also ...

Keywords: evolutionary algorithms, interactive search, interestingness measures, multiobjectiv optimization, rules mining

19 Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analys
Frizo Janssens, Wolfgang Glänzel, Bart De Moor
August 2007 KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledg discovery and data mining
Publisher: ACM
Full text available: pdf(865.75 KB) mov(18:26 MIN) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 38, Downloads (12 Months): 380, Citation Count: 0

To unravel the concept structure and dynamics of the bioinformatics field, we analyze a set of 7 publications from the Web of Science and MEDLINE databases, publication years 1981-2004. Fo delineating this complex, interdisciplinary field, a novel ...

Keywords: cluster chains, fisher's inverse chi-square method

20 A survey of Web metrics
Devanshu Dhyani, Wee Keong Ng, Sourav S. Bhowmick
December 2002 ACM Computing Surveys (CSUR), Volume 34 Issue 4
Publisher: ACM
Full text available: pdf(289.28 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 71, Downloads (12 Months): 679, Citation Count: 13

The unabated growth and increasing significance of the World Wide Web has resulted in a flurry research activity to improve its capacity for serving information more effectively. But at the hea these efforts lie implicit assumptions about "quality" ...

Keywords: Information theoretic, PageRank, Web graph, Web metrics, Web page similarity, qu metrics

Results 1 - 20 of 58 Result page: 1 2 3 next >>

Terms of Usage Privacy Policy Code of Ethics Contact Us